Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(8):209-217

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

A univariate analysis of molecular properties and inhibitory activity of dihydrothiophenones against dihydroorotate dehydrogenase of malaria parasite

Liza T. Billones[1] and Junie B. Billones[2,3*]

[1]*Department of Epidemiology and Biostatistics, College of Public Health*
[2]*Department of Physical Sciences and Mathematics, College of Arts and Sciences*
[3]*Institute of Pharmaceutical Sciences, National Institutes of Health*
University of the Philippines, Taft Avenue, Ermita, Manila 1000, Philippines

---

ABSTRACT

The recent discovery of dihydrothiophenone derivatives as P. falciparum dihydroorotate dehydrogenase (PfDHODH) inhibitors spurs quantitative examination of the relationship between the properties of these compounds and the observed antimalarial activity. Quantitative Structure-Activity Relationship (QSAR) study was carried out on dihydrothiophenones as inhibitors of PfDHODH in PfDd2 strain using multilinear regression analysis.The inhibitory activity was shown to be a function of eight DRAGON-type descriptors, namely, HATS7p, Hy, Mor17e, RDF145m, G1u, HATS8v, H5e, and Mor22m.The model indicates that an electronically dense molecule with highly electronegative atoms and less number of hydrophilic groups tend to be a potentantiplasmodial agent.

Keywords: QSAR, dihydrothiophenone, malaria, dihydroorotate dehydrogenase (DHODH), multiple linear regression (MLR), leave-one-out (LOO), leave-group-out (LGO), univariate analysis

---

INTRODUCTION

Malaria is a disease caused by the infection of *Plasmodium* parasites, which require two hosts in its life cycle: a mosquito vector and a vertebrate host. Among the five *Plasmodium* species that caused malaria, *P. falciparum* is responsible for the highest death rate and complications [1].According to World Health Organization (WHO), there were an estimated 213 million cases of malaria, and 655,000 malaria deaths worldwide in 2011 [1]. An estimated 3.3 billion people worldwide were at risk of malaria in the same year.

Although there are existing drugs against malaria, the emergence of drug-resistant strains of *Plasmodium*species has posed a serious health problem [2-4]. Resistance to traditional antimalarial drugs like chloroquine [5,6], sulfadoxine-pyrimethamine [3], and mefloquine has been a major concern [7].Moreover, recent reports by Hynes and co-workers [8,9]on artemisinin-based therapy have showed increased parasite clearance times with these agents signifying development of resistance. In fact, cases of artemisinin resistant malaria were already reported in at least two countries in Asia [1,10,11].

Needless to say, there is apressing need for new classes of effective antimalarial agents. In this light, the WHO Tropical Diseases Research (WHO-TDR) phenotypically screened around 5000 compoundsand identified a pyrrolone with remarkable activity against *P. falciparum.* Further SAR studies on pyrrolone-decorated compounds have demonstrated potential therapeutic application of these compounds against malaria [12]. However, more effort must be continually exerted in identifyingkey biochemical processesin host parasites and developing new drugs

against putative targets to combatdrug resistant malaria [13-15].

One biochemical pathway that is crucial for the survival of the parasite is thepyrimidine-based biosynthesis, an essential process for nucleotide production and cell proliferation [16,17].The key enzyme in this pathway is dihydroorotate dehydrogenase (DHODH), which mediates the fourth and rate-determining step in pyrimidine biosynthesis [18]. DHODH catalyzes the conversion of dihydroorotate (DHO) to orotate (ORO) [18,19].Since *P. falciparum* could not produce pyrimidine bases from other pathways, its dihydroorotate dehydrogenase (PfDHODH) has been regarded as an attractive drug target for antimalarial agents [20,21].

Recently, Xu and coworkers prepared a series of dihydrothiophenone derivatives and demonstrated the *in vitro* inhibitory ability of these compounds against DHODH, as well as the chloroquine- sensitive (Pf3D7) and resistant(*Pf*Dd2) strains [22]. With these valuable structure-activity data at hand, it is very encouraging to establish a quantitative model of antiplasmodial activity of these dihydrothiophenone variants. In this work, we performed a univariate analysis to derive a QSAR model based on DRAGON®-type molecular descriptors [23] in order to uncover the relevant molecular parameters that dictate the observed antimalarial activity of dihydrothiophenones.

EXPERIMENTAL SECTION

The biological data (IC_{50}) of dihydrothiophenone derivatives included in this study were obtained from literature [22]. IC_{50} represents the concentration of the compound (in μM) that affords 50% of the desired inhibitory activity. The 3D structures of all compounds were obtained using online program CORINA (http://www.molecular-networks.com).The molecular electrostatic potential maps were generated based on structures optimized at semi-empirical AM1 level using Spartan '14 ® (Wavefunction, Inc.) software. The molecular descriptors were calculated using the online program E-Dragon developed by the Milano Chemometrics and Todeschini QSAR Research Group(http://www.vcclab.org/lab/edragon/). Over 1200 classical descriptors were calculated that included constitutional (0D) properties, 1D descriptors (*i.e.* functional groups, atom centered fragments, information and properties descriptors, 2D descriptors (*i.e.* topological, molecular walk counts, Burden eigenvalues, eigenvalue based indices,topological charge indices, connectivity, edge adjacency and 2D autocorrelation descriptors, and 3D descriptors namely, charge, Randic molecular profiles, geometry, RDF, 3D-MoRSE, WHIM, and GETAWAY descriptors [23].

To establish a predictive structure–activity relationship, a multiple linear regression (MLR) equation or the QSAR model was obtained using the forward stepping protocol [24]in SPSS® version 20, which ran on Mac OS 10.8 system. A linear function with parameters α and β_i was generated that relates the *k*independent variables or descriptors (X_i) to the response variable *Y*in the form:

$$E(Y|X) = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k \qquad \text{(Equation 1)}$$

The quality of thefitted equation was initially evaluated by calculating the squared correlation coefficient(r^2), which indicates the proportion of the variation in the dependent variable that is explained by the regression equation [25]. Additionally, the multicollinearity among the predictors was examined using the bivariate correlation protocol in SPSS®.

Subsequently, the QSAR model was cross-validated using Leave-One-Out (LOO) and Leave-Group-Out (LGO) methods. In LOO technique [26], a single data value was removed from the dataset, then a new equation was derived based on remaining $n - 1$ dataset, and that equation was employed to predict the value of the datum that had been omitted.This process was repeated for every *y*valuein the dataset. In the LGO approach [27], a group of 8 compounds (20% of *n*) called test set was removed at each instance and a model was generated based on $n - 8$ dataset, also called as training set. After refitting, the *y*values for the excluded compounds were calculated using the fitted model. This process was repeated until all the *y* values have been calculated. Four more rounds of calculations were carried out so that the predicted activity of each compound is an average of five values.

The statistical validity of the model was assessedbased on cross-validated r^2, commonly known as q^2. While the r^2measures the goodness-of-fit, the q^2 measures the goodness of prediction [25]. The q^2 values for the LOO- and LGO-validated model were calculated from the model PRESS(prediction error sum of the squares) according to equations 2 and 3, respectively.

$$q^2_{LOO} = 1 - \frac{PRESS}{\sum_{i=1}^{N}(y_i - \langle y \rangle)^2}; \quad PRESS = \sum_{i=1}^{N}\left(y_i - y_{pred,i}\right)^2 \qquad \text{(Equation 2)}$$

$$q^2_{LGO} = 1 - \frac{PRESS}{\sum_{i=1}^{test}(y_i - \langle y_{train}\rangle)^2}; \quad PRESS = \sum_{i=1}^{test}\left(y_i - y_{pred,i}\right)^2 \qquad \text{(Equation 3)}$$

RESULTS AND DISCUSSION

Most drug discoveries today are outcome of an iterative cyclic, three-stage process that includes design, synthesis, and evaluation [25]. A routine approach in the design stage is the generation of QSAR models, which relate the observed bioactivity to the molecular properties. Such models are mathematical in nature and are constructed through application of appropriate statistical methods. One commonly employed statistical technique in QSAR studies is multiple linear regression (MLR), a supervised univariate method of analysis. MLR seeks to model the relationship between the independent variables (molecular descriptors) and the dependent variable (*i.e.* bioactivity)by fitting a linear equation to observed data.

In this study, the text file of molecular descriptors obtained from E-Dragon were exported to Microsoft Excel® and the resulting organized data were imported by SPSS® to act as the independent variables. Since high correlation has been demonstrated between the inhibitory activity against PfDHODH and antimalarial potency in both Pf3D7 (r = 0.87) and PfDd2 (r = 0.86) cells [22], only the observed half-maximal inhibitory concentration (IC_{50}) against the PfDd2 strain was chosen to serve as the dependent variable. After removing the entries with high and indefiniteIC_{50} values, only 39 dihydrothiophenones/dihydrofuranones were left and included in the MLR analysis. With this sample size (n = 39), the rule-of-thumb in model development requires that the MLR equation should contain no more than 8 descriptors (*i.e.* 5 samples per 1 independent variable) [25,28].Thus, an 8-descriptor QSAR model of PfDd2 activity was generated by the use of forward stepping regression method (equation 4).

$$\mathbf{IC_{50} = 53.341}\boldsymbol{HATS7p} \mathbf{+ 5.644}\boldsymbol{Hy} \mathbf{- 3.450}\boldsymbol{Mor17e} \mathbf{+ 6.210}\boldsymbol{RDF145m} \mathbf{+ 31.941}\boldsymbol{G1u} \mathbf{- 51.306}\boldsymbol{HATS8v} \mathbf{- 4.873}\boldsymbol{H5e} \mathbf{+ 2.135}\boldsymbol{Mor22m} \mathbf{- 4.714}$$

n = 39, r^2 = 0.910, F = 37.885, q^2_{LOO} = 0.831, q^2_{LGO} = 0.849 (Equation 4)

In forward stepping regression, the descriptor that provides the greatest contribution to the variation in the response variable is included first in the MLR model. The other descriptors were added next in order of decreasing importance, that is, according to their ability to explain the variability in the regressand.Equation 4 shows that *HATS7p*, a 3D GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) [29] descriptor, has the greatest contribution to the PfDd2 activity of the compounds under study. It single-handedly explains one-third (r^2 = 0.33) of the variation in PfDd2 activity. The positive coefficient of *HATS7p* indicates that a lower value for this predictor enhances antiplasmodial activity. *HATS7p* denotes leverage-weighted autocorrelation of lag 7, weighted by polarizability. It is based on a geometric distance matrix *H* and takes the form:

$$HATSkw = \sum_{i=1}^{n}\sum_{j=1}^{n}(w_i h_{ii})(w_j h_{jj})\delta(k; d_{ij}) \qquad \text{(Equation 5)}$$

where *w*, in this case, is a measure of polarizability, h_{ii} and h_{ij} are diagonal entries corresponding to the atoms *i* and *j* in the *H* Matrix, and $\delta(k, d_{ij})$ is Kronecker delta and equals unity if the ij^{th} entry in the Topological Level Matrix is equal to *k*, and zero otherwise [29]. Equation 5 implies that bydecreasing the polarizability (*w*) of the inner atoms and the overall size of the molecule the IC_{50} value would decrease (*i.e.* enhanced potency).

When the empirical index *Hy* was combined with *HATS7p*, more than half (r^2 = 0.55) of the variation in activity was accounted for. *Hy* is expressed largely in terms of number of hydrophilic groups (-OH, SH, NH) in the molecule. The QSAR model indicates that a molecule with fewer hydrophilic groups tend to be more potent against PfDd2 cells.Moreover, almost two-thirds of the variation in *y* (r^2 = 0.65) was explained by a 3-predictor combination that includes*Mor17e* as third descriptor.3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction), like *Mor17e*, are derived from IR spectra simulation using a generalized scattering function [30]. The MoRSE descriptor is defined as follows:

$$Mor(s,w) = \sum_{i=2}^{n}\sum_{j=1}^{i-1} w_i w_j sin(sr_{ij})/(sr_{ij}) \qquad \text{(Equation 6)}$$

where, *s* is a scatteringparameter, r_{ij} is the Euclidean distance between the atoms *i* and *j*, and *w* is an atomic property.*Mor17e* represents signal 17, weighted by atomic Sanderson electronegativities. The negative coefficient of *Mor17e* indicates that highlyelectronegative atoms in a molecule confer antiplasmodial activity.Taken together, a more potent antimalarial compound is electronically compact and contains highly electronegative atoms, albeit the hydrophilicity must be modulated.

A simple comparison of the molecular electrostatic potential mapsfor the most active compound (**50**, IC_{50} = 0.018

μM) and the least potent derivative (**48**, IC_{50} = 7.944 μM) illustrates our thesis (Figure 1). Obviously, compound **50** is less polarizable (*i.e.* more dense) and contains highly polar (*i.e.* more electronegative) hydrophilic functionalities compared to **48**. Our findings are consistent with Xu's [22] observation that a more compact ethoxycarbonyl is more favored over cyclopropylaminocarbonyl, which contains less electronegative N atom (relative to O) and bulkier alkyl group, at the 3-position of the pentagonal core. Furthermore, dihydrothiophenone ring is a less potent backbone compared to dihydrofuranone as indicated by the *HATS* and 3D-MoRSE predictors.

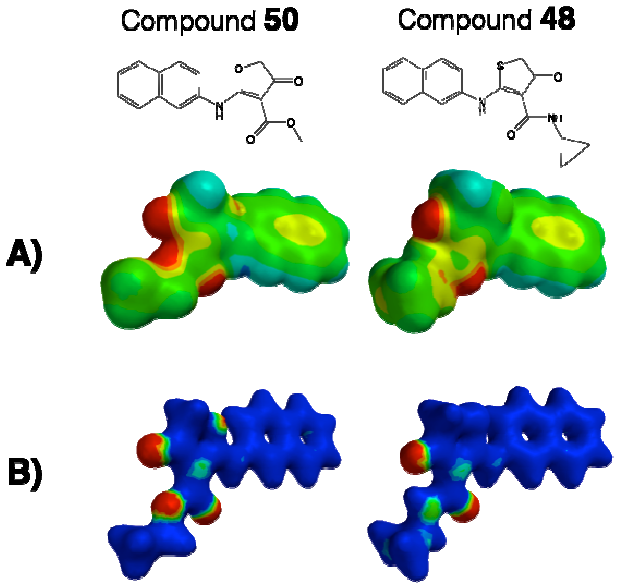
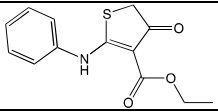
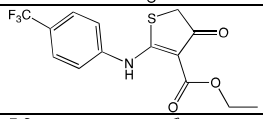
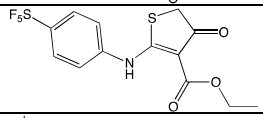
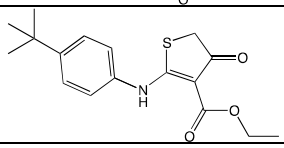


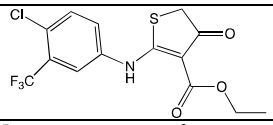
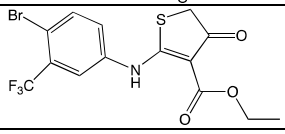
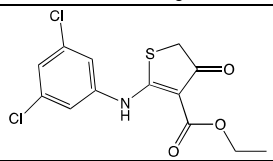
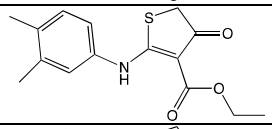
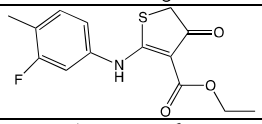
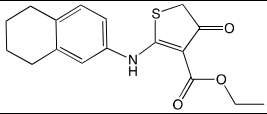
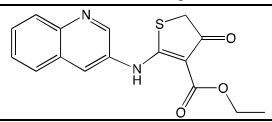
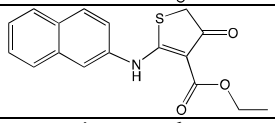
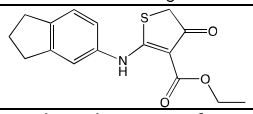
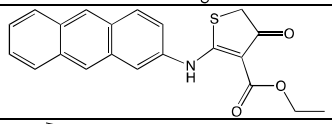
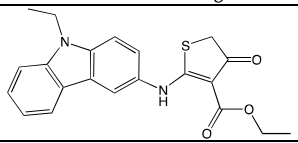
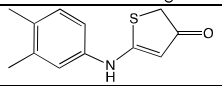
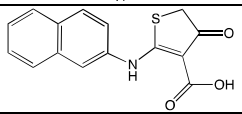
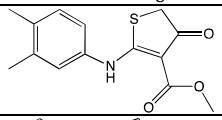
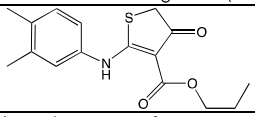
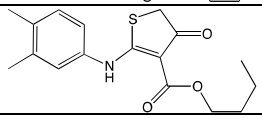
Figure 1.Molecular electrostatic potential map of 50 (column 1) and 48 (column 2). A) Electrostatic potential mapped onto molecular density isosurface, B) Electrostatic potential mapped onto bond density isosurface. The red spots locate the most negative portions in the molecule

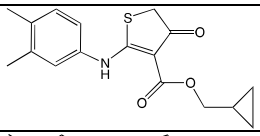
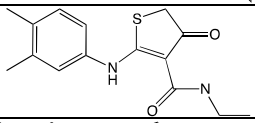
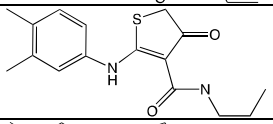
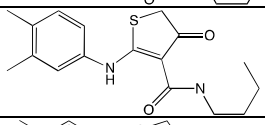
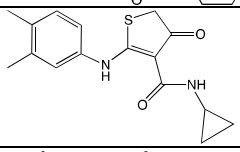
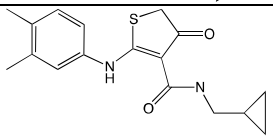
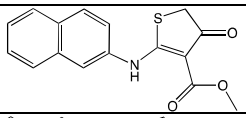
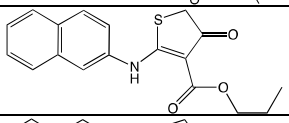
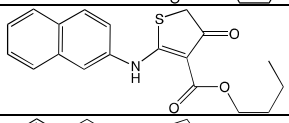
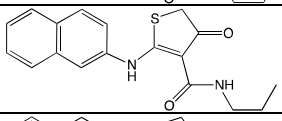
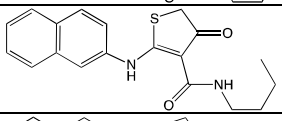
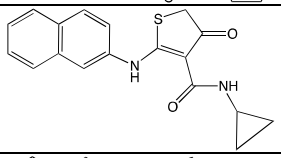
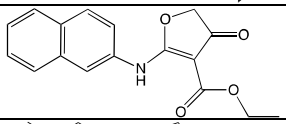
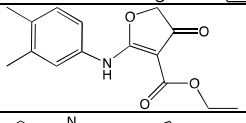
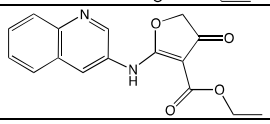
Interestingly, the next five descriptors addedto the model were all 3D parameters: *RDF145m* (radial distribution function descriptor), *G1u* (WHIM (Weighted Holistic Invariant Molecular) descriptor [31]), *HATS8v* and *H5e* (GETAWAY descriptors), and *Mor22m* (3D-MoRSE descriptor). In combination, these descriptors furnished roughly 26% additional contribution to the variation in the observed PfDd2 activity. The coefficients of these descriptors also suggest that a less massive thiophenone consisting of highly electronegative atoms is likely to exhibit greater antimalarial activity.

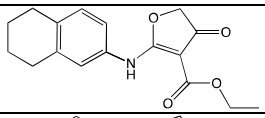
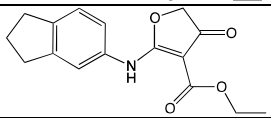
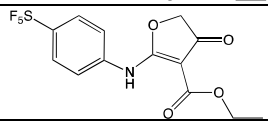
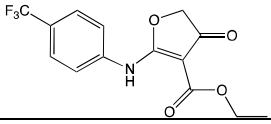
To evaluate the ability of the model for predicting IC_{50} values for a set of molecules, leave-one-out (LOO) and leave-group-out (LGO) cross validation procedures were carried out (*vide supra*). The experimental and predicted inhibitory activities of dihydrothiophenone derivatives against PfDd2 strain are presented in Table 1.

Table 1.Experimental [22]and Calculated IC_{50} values (μM) for the inhibitory action of dihydrothiophene derivatives against chloroquine-resistant strain of *P. falciparum* (PfDd2)

No.	Lit ID	Structure	Expt'l IC_{50}against *Pf*Dd2	Calc'd IC_{50}(LOO)	Calc'd IC_{50}(LGO)
1	**12**		6.556± 0.889	6.088	6.035
2	**13**		1.040± 0.146	0.997	1.080
3	**14**		0.697± 0.101	0.878	0.907
4	**15**		0.767± 0.367	-0.083	0.195

5	**16**		3.147± 1.018	2.120	1.983
6	**17**		5.539± 0.686	6.249	6.298
7	**19**		5.996± 0.636	4.916	5.264
8	**20**		1.317± 0.152	1.760	1.722
9	**21**		3.730 ± 0.915	3.054	3.200
10	**23**		0.315± 0.040	1.016	1.100
11	**24**		0.290± 0.030	1.366	1.509
12	**25**		0.057± 0.004	1.670	1.749
13	**26**		1.573± 0.192	2.760	2.789
14	**27**		1.127± 0.202	1.082	1.209
15	**28**		1.513± 0.864	1.452	1.515
16	**29**		2.498± 2.730	2.040	1.923
17	**31**		1.049± 0.079	1.619	1.305
18	**32**		1.267± 1.270	0.849	0.849
19	**33**		1.689± 1.665	0.692	0.628
20	**34**		0.889± 0.749	1.022	0.744

21	**35**		2.632± 0.961	2.615	2.489
22	**37**		6.625± 0.128	4.895	4.984
23	**38**		3.927± 0.775	3.532	3.486
24	**39**		2.778± 0.181	2.441	2.555
25	**40**		5.532± 0.992	6.957	7.249
26	**41**		3.500 ± 0.845	4.376	4.684
27	**42**		0.400 ± 0.265	0.382	0.193
28	**43**		1.660 ± 1.863	1.080	1.050
29	**44**		5.228± 1.017	3.071	3.565
30	**46**		3.547± 2.908	3.091	3.127
31	**47**		4.032± 0.025	5.757	5.303
32	**48**		7.944± 0.539	6.964	7.337
33	**50**		0.018± 0.002	0.002	-0.105
34	**51**		0.486± 0.012	0.707	0.636
35	**52**		0.517± 0.168	-0.615	-0.665

36	**53**		0.076± 0.008	-0.520	-0.626
37	**54**		0.373± 0.105	1.045	0.900
38	**55**		1.063± 0.414	1.214	1.063
39	**56**		0.867± 0.179	2.068	1.960

The plot of LOO-predicted IC_{50} values versus the observed activities (Figure 1) clearly demonstrates the predictive power of the QSAR model.The cross-validated squared correlation coefficient, q^2, of 0.83 indicates highly remarkable predictive ability, the normally recommended cut off being only 0.3 for a model to be considered statistically sound [32]. The q^2 value even surpassed the commonly considered acceptable value of 0.60 in QSAR studies [33]. Moreover, the scatter plot of the studentized residual (Figure 3) displays random distribution of errors around zero and shows no distinct pattern (r^2 = 0.0001) and outliers (t<±3.0).

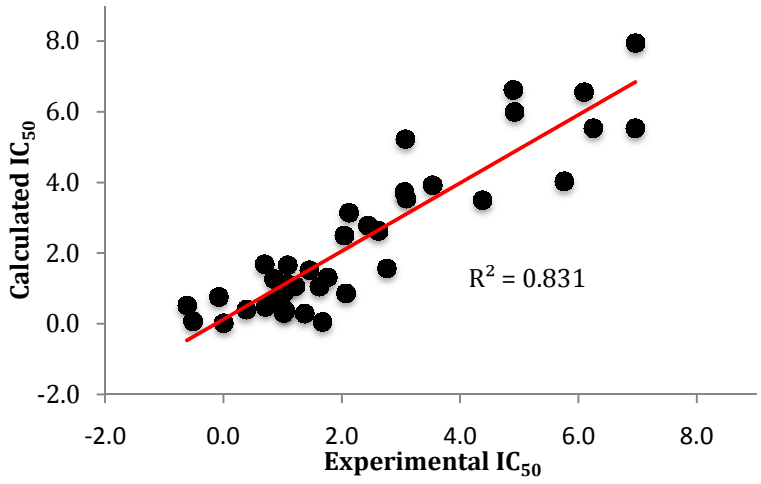


Figure 2. Calculated IC_{50} values by the use of Leave-One-Out (LOO) cross-validation approach vs. experimental IC_{50} data (µM)

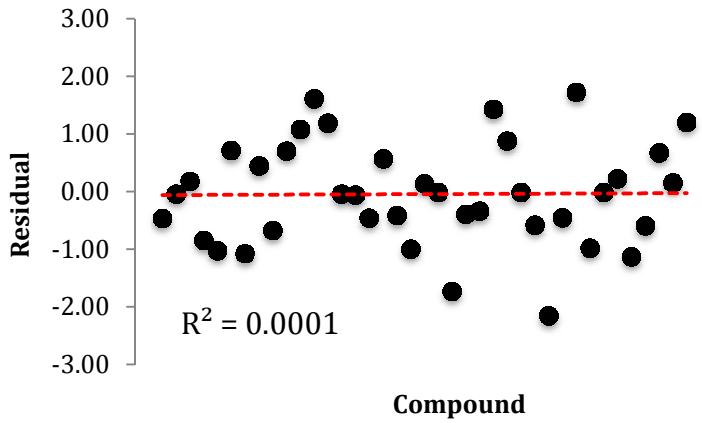


Figure 3. Studentized residual based on Leave-One-Out method

The results of LGO cross validation was closely consistent with the LOO method, the obtained q^2 value of 0.85 being slightly better than that of the latter (Figure 4). These results tend to corroborate the contention of Maw and

Hall [34] on the applicability of LGO technique for small data sets. The highly satisfactory statistics (r^2, q^2_{LOO}, q^2_{LGO}, absence of outlier and distinct trend in residuals) demonstrate the robustness of the QSAR model constructed from E-Dragon type descriptors.

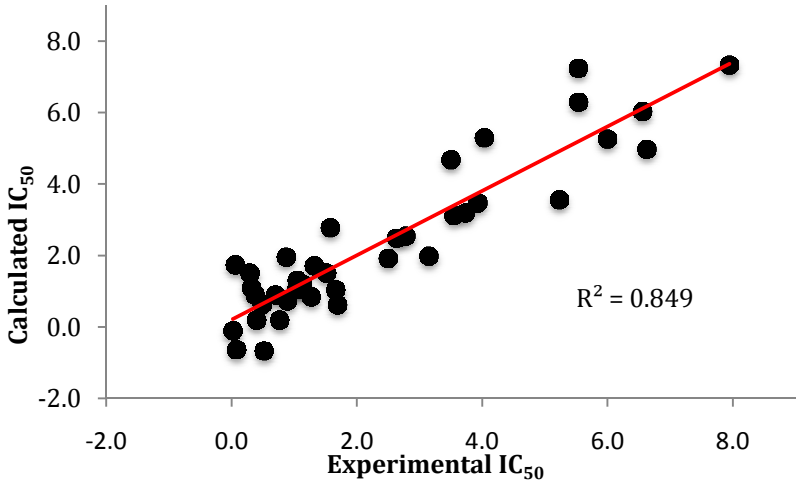


Figure 4. Calculated IC_{50} values by the use of Leave-Group-Out (LGO) cross-validation approach vs. experimental IC_{50} data (μM)

CONCLUSION

Quantitative structure-activity relationship (QSAR) study has been performed on dihydrothiophenone derivativeswith antiplasmodial activity. In particular, multilinear regressionanalysis was performed on the dataset composed of over 1200 E-Dragon type molecular descriptors, which served as independent variables, and inhibitory activity (IC_{50}) of 39 compounds against PfDd2 strain of malaria parasite that served as dependent variable. The 8-variable MLR model unveils that, aside from 1D hydrophilicity descriptor (*Hy*), the 3D parameters namely, two GETAWAY, two 3D-MoRSE,one WHIM, and one RDF descriptors are essential indicators for antiplasmodial activity. The model points to a relatively lighter and less polarizable molecule containing highly electronegative atoms but with less number of hydrophilic groups for greater potency against malarial parasite *P. falciparum*. These instructive results encourage the development of next generation antimalarial agents based on dihydrofuranone backbone.

Acknowledgement
JBB appreciates the *Gawad Sentenaryo* (Centennial Award) 2013 given by UP Manila and the computing facilities of the University of the Philippines EIDR Program (OVPAA EIDR 12-001-121102).

REFERENCES

[1] World Health Organization (WHO). World Malaria Report **2012**; WHO Press: Geneva, 2013. http://www.who.int/malaria/publications/world_malaria_report_2012. (Accessed November 28, 2013).
[2] TS Wu;AK Nagle; A Chatterjee.*Curr. Med. Chem*.,**2011**,18(6), 853−871.
[3] JH McKie; KT Douglas; C Chan C.;*et al*. *J. Med. Chem.*, **1998**,41(9), 1367−1370.
[4] JC Craft.*Curr. Opin. Microbiol.*, **2008**,11(5), 428−433.
[5] PG Bray; RE Martin; L Tilley;*et al.Mol. Microbiol.*,**2005**, 56(2), 323−333.
[6] JE Hyde. *Microbes Infect.*, **2002**, 4(2), 165−174.
[7] E Badawey; SM Rida; FSG Soliman; T Kappe.*J. Heterocyclic Chem*., **1989**, 26, 405−408.
[8] RK Haynes; B Fugmann; J Stetter;*et al*. *Angew. Chem. Int. Ed.*,**2006**, 45, 2082−2088.
[9] RK Haynes; WY Ho; HW Chan;*et al*. *Angew. Chem. Int. Ed*.,**2004**, 43, 1381−1385.
[10] AM Dondorp; S Yeung; L White;*et al.Nat. Rev.Microbiol.*, **2010**,8(4), 272−280.
[11] AM Dondorp; F Nosten; P Yi;*et al.N. Engl. J. Med*., **2009**,361(5), 455−467.
[12] D Murugesan;A Mital;M Kaiser;*et al.J. Med. Chem.*, **2013**, 56, 7911–2990.
[13] H Li; J Huang; L Chen;*et al.J. Med. Chem.*, **2009**, 52(15), 4936−4940.
[14] M Schlitzer.*ChemMedChem*., **2007**, 2, 944−986.
[15] MA Biamonte; J Wanner; KG Le Roch.*Bioorg. Med. Chem. Lett*., **2013**, 23, 2829−2843.
[16] M Löffler; LD Fairbanks; E Zameitat; AM Marinaki; HA Simmonds. *TrendsMol. Med.*, **2005**,11(9), 430−437.
[17] P Reyes; PK Rathod; DJ Sanchez;*et al*. *Mol. Biochem. Parasit.*, **1982**,5(5), 275−290.
[18] RL Fagan; MN Nelson; PM Pagano; BA Palfey. *Biochemistry*,**2006**,45(50),14926−14932.

[19] NA Malmquist; R Gujjar; PK Rathod; MA Phillips. *Biochemistry*, **2008**,47(8), 2466−2475.
[20] JM Coteron; M Marco; J Esquivias;*et al.J. Med. Chem.*,**2011**, 54(15), 5540−5561.
[21] H Munier-Lehmann; PO Vidalain; F Tangy; YL Janin. *J. Med. Chem.*, **2013**, 56(8), 3148−3167.
[22] M Xu;J Zhu;Y Diao;*et al.J. Med. Chem*., **2013**, 56,2975–7924.
[23] R Todeschini; V Consonni.Molecular Descriptors for Chemoinformatics, WILEY-VCH,Weinheim (Germany),**2009**, 1257.
[24] MA Efroymson. Multiple Regression Analysis. In: A Ralston;HS Wilf, Eds. Mathematical Methods for Digital Computers, New York: Wiley, **1960**, 191–203.
[25] AR Leach;VJ Gillet. Computational Models. In: An introduction to Cheminformatics, Revised Edition, Springer, The Netherlands, **2007**, 75–97.
[26] G Gong. *J Am. Stat. Assoc.*,**1986**, 81(393), 108–113.
[27] DW Osten.*Chemom.* **1988**, 2, 39−48.
[28] BG Tabachnick;LS Fidell. Multiple Regression. In: Using Multivariate Statistics. 5th Edition, London: Pearson/Allyn & Bacon, **2007**, 123.
[29] V Consonni;R Todeschini;M Pavan.*J. Chem. Inf. Comput. Sci.*, **2002**, 42, 682–692.
[30] LJ Soltzberg;CL Wilkins. *J. Am. Chem. Soc.*, **1977**, 99, 439–443.
[31] R Todeschini;M Lasagni;E Marengo. *Theory J. Chemometrics*, **1994**, 8, 263–273.
[32] U Norinder. Calculated Molecular Properties and Multivariate Statistical Analysis. In: H van de Waterbeemd;B Testa, Eds. Drug Bioavailability. 2nd Edition. Vol. 40. Weinheim: Wiley-VCH Verlag GmbH, **2009**, 399.
[33] S Wold.*Quant. Struct. Act. Relat.*, **1991**, 10(3),191–193.
[34] HH Maw; LH Hall.*J. Chem. Inf. Comput. Sci*.,**2001**, 41,1248–1254.